INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE
DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT
SDC -- A BURROUGHS COMPANY

ARPA ORDER NUMBER: 5262
PROGRAM CODE NO. NR 049-602 dated 10 August 1984 (433)
CONTRACTOR: System Development Corporation CONTRACT AMOUNT: $683,105
CONTRACT NO: N00014-85-C-0012
EFFECTIVE DATE OF CONTRACT: 4/29/85 EXPIRATION DATE OF CONTRACT: 4/29/87
PRINCIPAL INVESTIGATOR: Dr. Lynette Hirschman PHONE NO. (215) 648-7554

SHORT TITLE OF WORK: DARPA Natural Language Understanding Program

REPORTING PERIOD: 11/1/85 - 1/31/86

1. Description of Progress

1.1. Grammar

A translator which translates hnl' grammar rules to Prolog has been implemented which allows grammar rules to be compiled. It is now possihle for translated and interpreted grammar rules to be run intermingled during development.

The grammar has been extended to handle structures containing wh-gaps, including overt wh-questions, such as *What did the field engineer inspect?*; relative clauses, such as *The disk which failed was bad.* and indirect questions, as *He asked what she was repairing.* The grammar can also handle the interaction between wh-structures and conjunction. This interaction can be complex because both constructions involve gaps. For example, in *The disk which he repaired and installed has failed,* (which our system can parse) the objects of *repair* and *install* as well as the subject of *install* are syntactic gaps.

A Prolog version of the representation developed at NYU to mediate between the syntactic parse and semantics has been implemented and is currently being tested.

1.2. Semantics

Semantic analysis procedures for one category of fragmentary sentences, those with missing subjects, have been implemented. Missing subjects are treated with the same mechanism as that used for pronouns. That is, a list of possible referents, the **FocusList** is maintained and the first candidate satisfying the verb selectional restrictions is chosen as the referent.

A treatment of *non-specific* noun phrases such as *a disk drive* in *fe ordered a disk drive* was implemented. This noun phrase does not refer to any particular disk drive, as does a *specific* noun phrase, as in *fe inspected a disk drive*, and this difference must be represented for the semantic representation to be accurate. Handling this requires close cooperation between semantics and reference resolution, because what determines the specific vs. non-specific interpretations of the noun phrases is the type of the verb. Verbs such as *order* are represented as having non-specific objects. When reference resolution is called for a noun phrase object of one of these verbs, the information that the noun phrase is non-specific is passed in so that reference resolution knows not to look for a specific referent. Instead, it leaves the referent as an uninstatiated variable, which may be filled in at a later point in the text.

A treatment of temporal relationships among events mentioned in the CASREPS has been designed and is being implemented. The critical tasks for adequately representing temporal relations are first, to distinguish between states of affairs (e.g., faulty equipment) versus changes of state (e.g., equipment failures), and second, to establish the relative ordering of these states and state changes, or events. Work on the first task involves an analysis of the semantics of verbs and the interaction between verb semantics and tense. The component which will compute the verb-dependent interpretation of tense is a natural extension of the existing logic-based verb semantics. It is based on a modification of Palmer's semantic analysis of verbs in order to provide different logical representations for 'state' and distinguish aspectual classes of verbs.

Natural language texts generally provide enough information to generate only a partial ordering. In the current domain, information about simultaneity or succession is generally expressed by means of adverbial conjunctions and prepositions such as *after, before, during,* and *when.* The interpretation of such expressions depends on representing the time of the event or state expressed in one part of a sentence in relation to that expressed in another part of the same sentence. Thus, the time of the state/event following an adverbial expression like *before* serves as a temporary point of 'reference.' The notion of a temporary reference point is a simplification of Hirschman's work on the more complex temporal relations expressed in narrative text. The temporal information pertaining to the entire set of states and events mentioned in a descriptive text can be aggregated into a single discourse representation in terms of a single 'reference period.'

Implementation

To implement the time analysis, changes are being made to the semantic decompositions of verbs and nominalizations and to the clause semantics interpreter. Verbs and nominalizations which can refer to events will contain a **BECOME** predicate to distinguish them from stative predicates. Event predicates are a heterogeneous class, including both inchoatives and causatives. Sometimes a single morphological form can occur in all three aspectual classes, e.g., *disengaged.* The state predicate in *The air compressor is disengaged* would be represented as **disengaged(theme(X))**; in *The air compressor disengaged after the alarm*, the inchoative event would be represented as an embedding of a 'disengaged' state under the aspectual predicate **become**: **become(disengaged(theme(X))**; finally, the causative event expressed in *the operator disengaged the air compressor* would be represented with the inchoative event embedded under a **cause** predicate: **cause(agent(X),become(disengaged(theme(X))))**. The clause semantics interpreter is being upgraded with a time component to take advantage of the aspectual operators being added to the verb semantics.

The time component of the interpreter will be responsible for creating representations of states and events with appropriate time arguments. In order to compute whether a given clause expresses an event or a state, the time component of the interpreter will look at the lexical decomposition of the verb, to see if it contains a **become** predicate, and at the tense and aspect information from the sentence.

Given a sentence in the simple past tense with an 'event' verb (i.e., a verb whose decomposition includes the predicate **become**), the time component will first create an event representation with an associated **moment** argument; e.g., Sentence: *The air compressor disengaged after the alarm.*

Event Representation: **event(e1,**
'disengage',
become(disengaged(theme('air compressor'))),
moment(m1))

where **e1** is a pointer to the event and **m1** is a pointer to the moment at which that event occurred. Since the result of an event is that a new state of affairs has occurred, the time component will also create a corresponding state representation. Thus, after the event in which the air compressor becomes disengaged, it is in a state of being disengaged, and will remain so for an indefinite period of time:

state(s1,
'disengage',
disengaged(theme('air compressor')),
period(p1,after(m1,0)))

Here, **s1** and **p1** are pointers to the state and corresponding period of time. Note that the period of 'disengagement' **p1** is represented as beginning after the moment **m1** in which 'disengagement' occurred. The second argument of **after** shown above indicates that there is no time interval between the moment **m1** and the period **p1**. That is, **m1** is the onset of **p1**. The interval argument of the **after** predicate makes it possible to represent the meaning expressed in sentences like *The air compressor disengaged three minutes after the alarm sounded*, in which there is a lapse of time between the onset of one event ('sounding of alarm') and the onset of a second event ('disengagement').

In order to determine whether a sentence makes reference to a state or an event, one has to take into account not only the inherent meaning of the verb, represented in the semantic decompositions, but also the tense and aspect in which the verb appears in the sentence input. An 'event' verb in the simple past tense generally refers to an event. However, the same verb will refer to a state rather than an event if it occurs in the present tense or progressive aspect. The following sentences with the verbs *drop* and *decrease*, which have the same semantic decompositions, illustrate the effect of tense and aspect: *The air pressure dropped below 30 psig. Lube oil pressure decreases below alarm point after engagement. The lube oil pressure is dropping.* The first sentence, in the past tense, refers to an event, i.e., to an occurrence which is the onset to a new state. The second, in the present tense, refers to a state, i.e., to a period of time in which some event repeatedly occurs. The last sentence, in the progressive aspect, also refers to a state of affairs, one in which a particular activity is occurring over a period of time with an unspecified endpoint. One can infer, of course, that the period of time in which the pressure is dropping will have to end at some future moment; that moment would then be the onset of a new state of affairs, e.g., one in which the lube oil pressure is low.

1.3. Facilities

The natural language system has been ported to Symbolics Prolog 6.1. This port was delayed due to a number of bugs in Symbolics Prolog and the poor development environment provided for Prolog by Symbolics. A Quintus Prolog compatibility package is under development which will facilitate maintaining common code between our systems on the VAX and the Symbolics.

We are still awaiting receipt of Prolog on GFE.

The version of the natural language system on the VAX has been ported from Quintus Prolog 1.0 to Quintus Prolog 1.5.

2. Change in Key Personnel

Leslie Riley (M.S., University of Pennsylvania, 1985) started on November 4.

John Dowding (M.S.E. expected 1986) resumed full time education, at the University of Pennsylvania, and is now working part time at SDC. Rebecca Davis left SDC on December 6.

3. Summary of Substantive Information from Meetings and Conferences

3.1. Professional Meetings Attended

Linebarger, M., and Schiffman, R. attended the Annual Meeting of the Linguistic Society of America. Seattle, December, 1985.

3.2. SDC/NYU Meetings

SDC/NYU Meeting #6 (November 1, New York University)

Lynette Hirschman, Martha Palmer, Deborah Dahl, John Dowding, Marcia Linebarger, and Rebecca Schiffman attended a meeting at NYU with Ralph Grishman, Ngo Thanh Nhan, and Tomasz Ksiezyk. Palmer and Schiffman gave a presentation on semantics rules for the CASREP's. Ksiezyk presented a description of the starting air system. During the afternoon the meeting broke up into smaller groups to discuss specialized issues in the grammar (Nhan, Hirschman, and Linebarger), the intermediate representation (Grishman, Dahl, and Dowding), and semantics (Ksiezyk, Palmer, and Schiffman).

SDC/NYU Meeting #7 (January 24, SDC, Paoli, PA) Ralph Grishman, Tomasz Ksiezyk, Dimitri Turchin, and two other members of the NYU group came to Paoli to meet with Lynette Hirschman, Martha Palmer, Deborah Dahl, John Dowding, Marcia Linebarger, Leslie Riley, and Rebecca Schiffman. Grishman gave a presentation on current developments at NYU. Ksiezyk described a detailed domain model of the starting air compressors. Palmer demonstrated the SDC system and discussed verb decompositions. Ksiezyk demonstrated a graphic model of the SAC's on the Symbolics.

3.3. Symbolics Lisp User's Group

John Dowding and Leslie Riley attended the meeting of the Mid-Atlantic Division of the Symbolics Lisp User's Group at the University of Pennsylvania on January 9.

4. Problems Encountered and/or Anticipated

Prolog for the Government-furnished Symbolics machine has not yet arrived.

Problems remain with with the lack of development environment and debugging facilities in Symbolics Prolog.

5. Action Required by the Government

Prolog for the Government-furnished Symbolics machine.

6. Fiscal Status

(1) Amount currently provided on contract:
$339,728 (funded) $683,105 (contract value)

(2) Expenditures and commitments to date:
$219,103 (through January 31, 1986)

(3) Funds required to complete work:
$120,625 (Year 1) $464,002 (Yrs. 1-2)